ENHANCING THE SYNTHESIZABILITY OF DPO4 POLYMERASE

by

David Paul Judd

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Biochemistry

The University of Utah

August 2015

The University of Utah Graduate School

STATEMENT OF THESIS APPROVAL

The thesis of David Paul Judd

has been approved by the following supervisory committee members:

Michael S. Kay	, Chair	3/27/2015 Date Approved
Christopher P. Hill	, Member	3/26/2015 Date Approved
Brenda L. Bass	, Member	3/26/2015 Date Approved

and by Christopher P. Hill , Chair/Dean

of the Department of Biochemistry

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

The total chemical synthesis of 361-residue Dpo4 DNA polymerase (including N-terminal His-tag) is an ambitious undertaking that pushes the limits of solid-phase peptide synthesis (SPPS). In chemical protein synthesis projects, smaller peptide segments are synthesized using SPPS and these segments are ligated together via native chemical ligation (NCL) to produce a full-length protein. NCL is a chemoselective reaction that requires a C-terminal thioester and an N-terminal Cys residue to ligate peptide segments. These requirements pose a challenge to synthetic protein chemists because Cys residues are rare in most proteins and C-terminal thioester selection is crucial to the success of the project because some thioesters react quickly and efficiently while others are slow and produce very low yields. A strategically placed mutation, usually to optimize ligation sites, can make an impossible chemical protein synthesis project possible. However, mutations must be chosen carefully to minimize their impact on the activity of the protein.

In the case of Dpo4, we identified three positions where mutations would greatly simplify our chemical synthesis efforts. In order to safely make mutations, we used BLAST searches and crystal structure analysis to determine candidate residues. We recombinantly expressed mutated Dpo4 constructs and compared their activity against wild-type Dpo4. In this manuscript, we discuss the selection of the three mutated positions and the development of a PCR assay that allowed

us to quantify differences in thermostability between mutant Dpo4 constructs and wild-type Dpo4. This research highlights the importance of strategically placed mutations in chemical protein synthesis projects and also has the potential to impact the biotechnology field with the introduction of mirror-image PCR using chemically synthesized D-Dpo4 as the DNA polymerase.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank some very important people who contributed greatly to this project. First of all, I thank Michael Kay, my advisor, for providing me the opportunity and the resources to make progress on this project. I also thank Michael Jacobson for assisting me with all aspects of this project from experimental design to assistance with formatting figures for the written document. I am deeply grateful for their assistance throughout this process.

I would also like to thank Amanda Smith and Nico Szabo from the Kay lab for their advice on PCR-related aspects of the project and Mark Peterson for always being available to answer any chemistry-related questions that arose throughout the project. Two members of the Hill lab (Casey Hemmis and Erik Fredrickson) were also very helpful and had great advice regarding cloning, protein expression, and protein purification strategies.

CHAPTER 1

INTRODUCTION

History of Chemical Peptide Synthesis

Chemical peptide synthesis has a long and storied history stretching back to the early 1880s when Theodor Curtius successfully prepared an N-protected dipeptide benzoylglycylglycine. This dipeptide was prepared by treating the silver salt of glycine with benzoylchloride. The azide-coupling method discovered by Curtius and an acylchloride coupling method discovered by Emil Fischer were used for many years to chemically synthesize peptides. However, these methods, employing carboxyl protecting groups, were limited by near complete racemization of the final product. In contrast, an amino-protecting group, carboxybenzyl (Cbz), was developed by Max Bergmann (a former student of Emil Fischer) and Leonidas Zerwas, that increased the utility of peptide synthesis by greatly minimizing racemization (Kimmerlin 2005).

Until 1963, all peptide synthesis chemistry was done using solution-phase techniques that required each reaction to be isolated and purified before the addition of subsequent amino acids (Kimmerlin 2005). Bruce Merrifield then pioneered a method known as solid-phase peptide synthesis (SPPS) in which each reaction is linked to an insoluble bead (Merrifield 1963) (Kimmerlin 2005). After each coupling reaction, the solid support can be washed to remove any

excess amino acid or coupling reagent, deprotected, washed again, and then coupled with the next amino acid. Synthesis proceeds from C- to N-terminus, opposite from protein generation in a biological system. Figure 1A shows Bruce Merrifield's original synthesis scheme from his seminal paper in 1963, and Figure 1B shows the general methodology used for SPPS.

There have been many advances in the field of SPPS specifically relating to protecting groups and the chemistries involved in synthesizing peptides. Today, there are two main amino-protecting groups in use that differ in the chemistry involved to remove them to prepare for the next amino acid coupling step. First, *tert*-butyloxycarbonyl (Boc) is an acid-labile amino-protecting group that requires the use of concentrated trifluoroacetic acid (TFA) for removal (Kent 1988). Second, fluorenylmethyloxycarbonyl (Fmoc) is a base-labile group that is removed upon treatment with piperidine (Carpino 1972). In addition to these two main amino protecting groups, there are many other side chain protecting groups that prevent unwanted chemical reactions on side chains (Kimmerlin 2005). The advent of SPPS greatly stimulated the field of peptide synthesis by making the process much faster and simpler, eventually allowing for automation (Kent 1988).

In the Kay lab, we have chosen to use Fmoc SPPS because it is much safer than the Boc method. Boc SPPS methods require treatment with TFA for every deprotection step and treatment with HF, an even harsher acid, for cleavage from the resin (Kent 1988). We chose the Fmoc route to avoid dealing with pressurized TFA at every deprotection step. Fmoc SPPS only requires a small amount of TFA for cleavage from the resin at the end of the synthesis. Figure 2A shows the chemical structure of Fmoc-Cl, which is coupled to the

amino group of all twenty amino acids. Figure 2B depicts the deprotection reaction that takes place upon treatment with piperidine. After cleavage with TFA, which also removes the side chain protecting groups, the peptide is precipitated with ether and then purified using reverse-phase HPLC and analyzed using mass spectrometry.

Chemical Synthesis of Proteins

The study of proteins in biology has long been of interest both structurally and functionally. The most common way of obtaining protein samples is by use of recombinant DNA techniques where a gene of interest is cloned into an expression plasmid that is introduced into an organism, mainly bacteria or yeast. The protein is then expressed and purified from the host cell. Recombinant techniques work well for most proteins, but there are limitations such as the fact that some proteins are toxic or do not express well (Saida 2007). Additionally, it is possible but very difficult to express unnatural amino acids in a recombinant system (Xie 2005).

The major advantage of complete chemical synthesis of proteins is that the chemist has control over the chemical makeup of the protein. This allows the incorporation of unnatural amino acids such as D-amino acids and fluorescent reporter groups (Kimmerlin 2005). The limiting factors in chemical synthesis of peptides and proteins are the final yield and purity of the product. Not every deprotection or coupling reaction is 100% efficient, leading to losses in yield and purity with each amino acid incorporation step (Kimmerlin 2005). Due to decreasing yield with increasing peptide length, peptide synthesis is generally

limited to products of 50-70 amino acids long, though this limit is sequence dependent.

In order to synthesize longer peptides or proteins, it is necessary to combine peptide segments synthesized via SPPS in a ligation reaction known as native chemical ligation (NCL). NCL is a chemoselective reaction between a C-terminal thioester and an N-terminal Cys (1,2-aminothiol), forming a native amide bond (Dawson 1994) (Johnson 2006) (Malins 2013) (Figure 3). Since native Cys residues are not common in most proteins, synthetic peptide chemists have devised a way to use Ala residues as junctions for NCL. The junction Ala residue is synthesized as a Cys residue and, after NCL has been completed, a free-radical-based desulfurization reaction is employed which converts the Cys residue into an Ala residue (Dawson 2000) (Yan 2001). Using Ala residues as NCL junctions greatly increases the number of potential junction sites, but also adds complexity to protein synthesis. For example, if a native Cys residue needs to be maintained (rather than desulfurized) in the final protein, then it must be protected, typically with an acetamidomethyl (ACM) thiol-protecting group (Isidro-Llobet 2008). The ACM group then needs to be removed after the desulfurization reaction, a step that often results in significant loss of product (Harris 2007).

A key factor that influences the rate and success of NCL is the C-terminal thioester residue. Some residues complete the NCL reaction in less than 4 h, while others require more than 48 h (Hackeng 1999). Thus, strategic selection of ligation sites in a protein synthesis project is a very important consideration that will be discussed further in the results section.

Kay Lab Goals

An important application of chemical protein synthesis is in preparing mirror-image D-peptides for use as therapeutics, and the Kay lab has long been focused on the use of D-peptides as viral entry inhibitors. D-peptides are composed of D-amino acids and are mirror images of L-peptides (Figure 4). Unlike most small molecule inhibitors, peptides are capable of disrupting protein-protein interactions otherwise considered "undruggable" targets. The major disadvantage of peptide drugs is that they are rapidly broken down by proteases, hence requiring frequent and high dosing to be effective. D-peptides are not broken down by L-proteases and therefore have a much longer half-life in the body (Weinstock 2012). The Kay lab currently has a D-peptide HIV entry inhibitor that is in late preclinical trials (Francis 2012) (Welch 2010). Some of the same methodologies that yielded success against HIV are now being applied to the development of a D-peptide Ebola entry inhibitor (Clinton 2014).

Along with chemical synthesis of D-peptide therapeutics, we are also interested in engineering other ways to produce D-peptides and D-proteins, which would greatly simplify the process and also allow for large-scale production. Our goal is to produce an in vitro translation system for the production of D-proteins by synthesizing all components of the ribosome and assembling them into a functional mirror-image ribosome. With an assembled D-ribosome it should be possible to produce D-proteins when the system is provided with mirror-image RNA, mirror image tRNA's, mirror-image GTP and ATP (energy sources), and D-amino acids. Such a mirror image in vitro translation system would be a great step forward for the routine production of

large D-proteins. The long-term goal is to design a synthetic mirror-image organism that we have termed '*D. coli*', which would be capable of producing D-proteins in a large-scale format.

DapA Review (Weinstock 2014)

With the chemical synthesis of larger D-proteins becoming feasible (though still very difficult), it is now necessary to take a step back and contemplate other difficulties that may arise when working with large D-proteins. In cells, there are multiple systems set up to assist folding proteins as they are translated by the ribosome. Specifically, many larger proteins cannot fold into their active forms without the help of chaperones (Frydman 2001).

The folding of large D-proteins is therefore an area of great interest if the field of D-protein synthesis is to advance to larger targets. If L-chaperones are unable to fold D-proteins, then it would be necessary to synthesize and fold D-chaperones, which would greatly complicate these already difficult projects. With this in mind, the Kay lab set out to discover whether the *E. coli* GroEL/ES chaperone can fold a D-protein.

DapA (4-hydroxy-tetrahydrodipicolinate synthase) is an *E. coli* protein that relies upon GroEL/ES to fold into its active form (Kerner 2005). DapA catalyzes a key step in lysine and diaminopimelic acid biosynthesis. Interestingly, when GroEL/ES is not present, DapA is aggregated and degraded leading to cell death (Kerner 2005). Based on these data, the Kay lab chose DapA as its first large D-protein synthetic target.

In order for this project to be successful, the 312-residue DapA protein

would have to be synthesized in D. Then, both L- and D-DapA would be subjected to folding assays using L-GroEL/ES to determine the chaperone's stereospecificity. If correct folding was achieved, then functional assays could be performed to determine the activity of D-DapA compared to synthetic L-DapA and recombinant DapA.

On the path to synthesizing the 312-residue DapA, the sequence was analyzed to find all Cys and Ala residues, as these could be used as ligation sites for NCL. The protein was broken down into eight segments with each containing an N-terminal Cys or Ala. This strategy required 12 handling steps (ligations, desulfurizations, and ACM removals) and proved impossible due to cumulative sample losses. The assembly of the N-terminal half of the protein could be made much easier by eliminating the desulfurization reaction at the junction between segments 2 and 3. Based on this idea, a BLAST search was done comparing the 1,000 most-similar homologs of DapA, and it was found that 12% of these proteins had a Cys instead of an Ala at position 77. Using this information as well as crystal structure analysis, it was decided that the synthesis should go forward using the A77C mutation to simplify the assembly.

Another useful change to the synthesis strategy included combining segments 7 and 8 in the initial synthesis. Combining other peptide segments was also pursued, but it was not possible to synthesize and purify these larger peptides with acceptable quality and yield. These two changes, A77C and the combination of segments 7 and 8, removed four handling steps and their associated sample losses, enabling the successful assembly of the 312-residue DapA in both L- and D- forms. Figure 5 shows the sequence of DapA along with

the synthesis strategy and Figure 6 shows the HPLC and mass spectrometry data obtained on the final L- and D-DapA synthetic products.

Synthetic DapA was then subjected to folding experiments using GroEL/ES, and enzymatic activity assays were performed comparing recombinant wild-type protein to synthetic L and D-DapA. Figure 7 shows that L-GroEL/ES is able to refold D-DapA confirmed by activity assay normalized to recombinant DapA refolded by GroEL/ES.

The fact that GroEL/ES does not exhibit strict stereospecificity in folding its substrates is a major finding that has many implications for the future of D-protein synthesis. Given this result, many more D-protein targets are now potential candidates to be synthesized and further explored.

Using the knowledge that we gained from the DapA project, we are now attempting an even more ambitious synthetic protein project of the 352-residue DNA polymerase Dpo4. We plan to synthesize this archeal DNA polymerase in both L- and D- forms and validate its function using PCR and extension assays. The D- form of Dpo4 will also have practical utility by enabling mirror-image PCR.

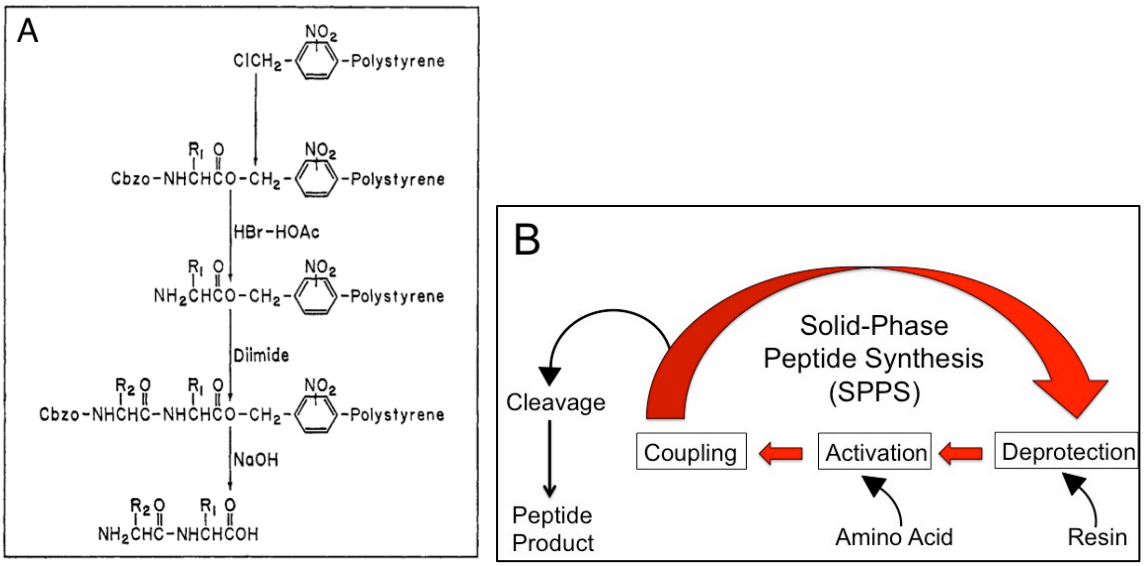


Figure 1: Solid-Phase Peptide Synthesis. (A) Bruce Merrifield's original depiction of SPPS using a polystyrene bead from his seminal paper (Merrifield 1963). (B) Diagram showing the process of SPPS. The general cycle progresses from deprotection of the amino terminus, wash, amino acid activation, coupling reaction, wash, then deprotection and repeat. When the synthesis is complete, the peptide is cleaved from the resin.

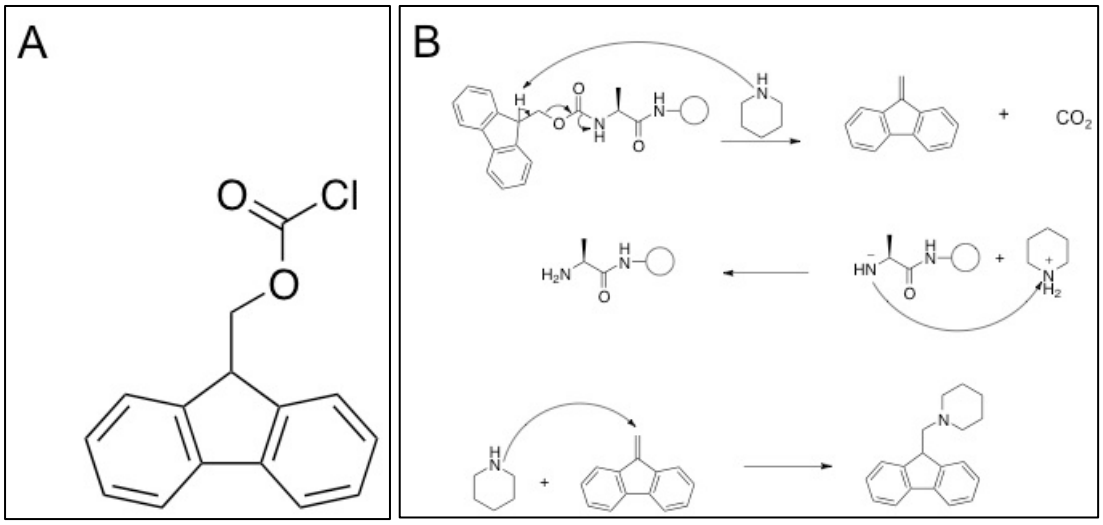


Figure 2: Fmoc-Cl and Fmoc deprotection scheme. (A) Chemical structure of Fmoc-Cl. Fmoc-Cl is attached to the amino terminus of all amino acids and serves as a protecting group preventing unwanted couplings. (B) Piperidine-catalyzed Fmoc deprotection of an amino acid on-resin. This deprotection step allows the next amino acid coupling step.

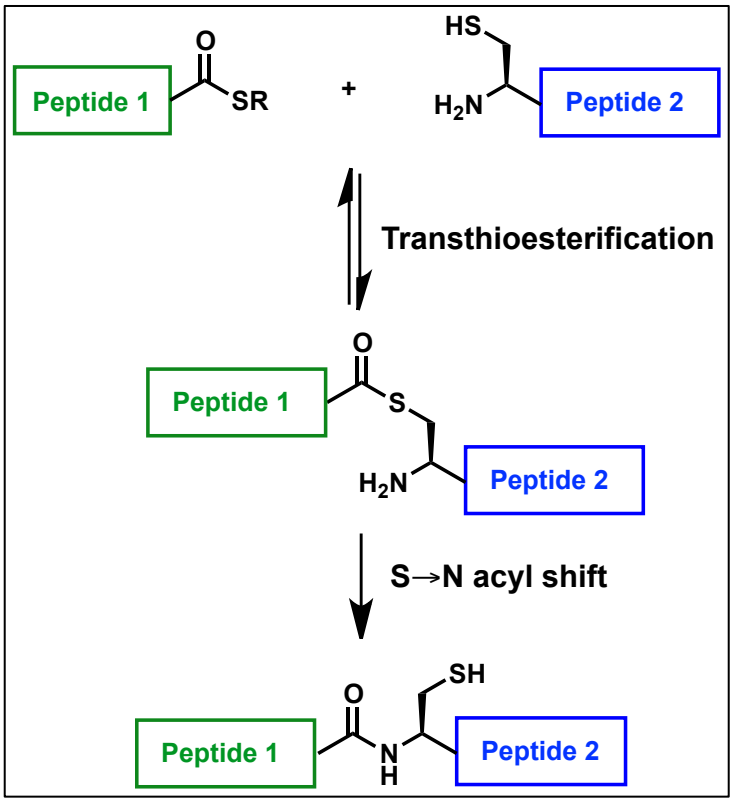


Figure 3: Mechanism of native chemical ligation. A reaction between a C-terminal thioester and the thiol group of an N-terminal Cys residue occurs and then undergoes an acyl shift leading to the formation of a native amide bond.

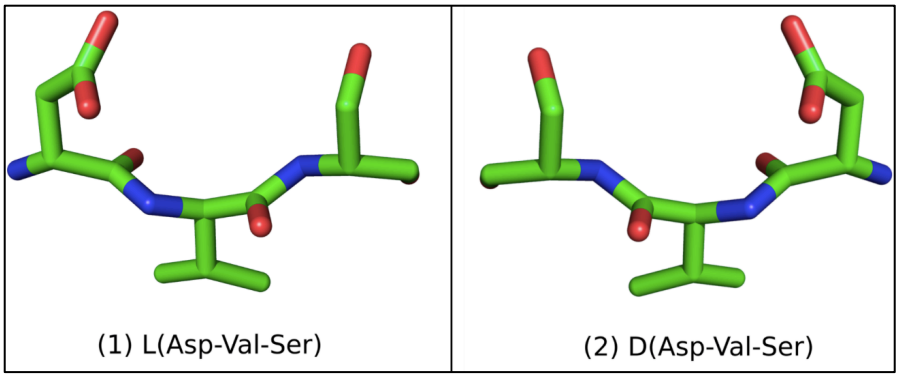


Figure 4: Mirror-image peptide sequence. (1) shows the sequence Asp-Val-Ser in L-form while (2) shows the same sequence in D-form. They are mirror images of each other. Image from (http://upload.wikimedia.org/wikipedia/commons/1/15/L-Peptide-D-PeptideMirrorImages.png).

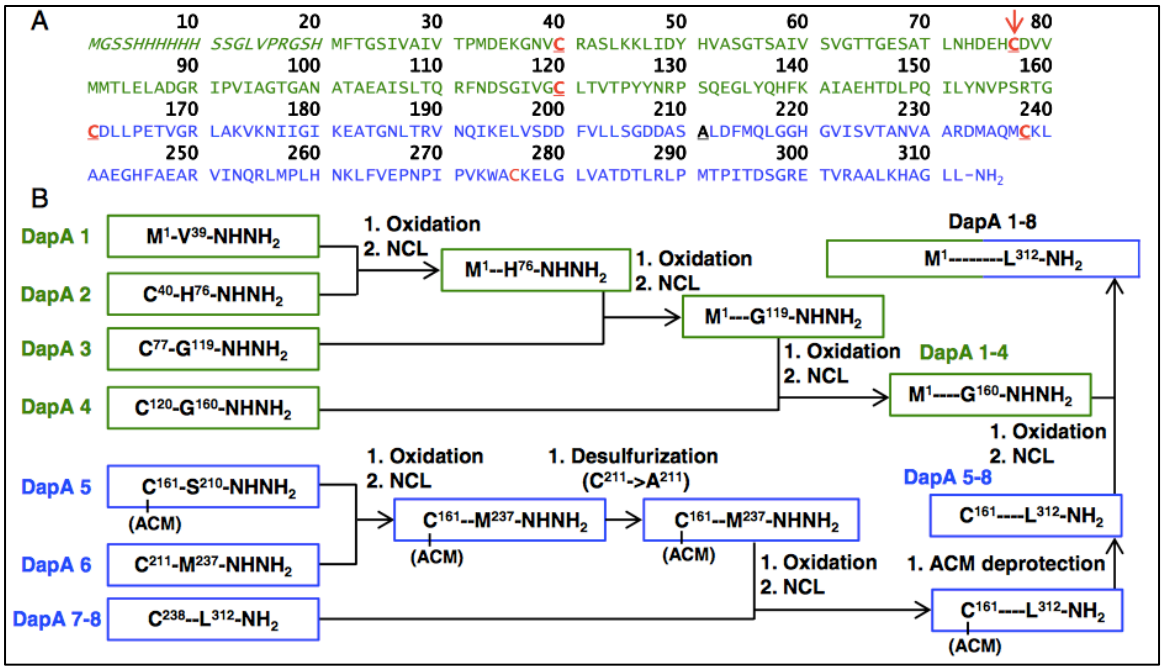


Figure 5: Synthesis strategy for 312-residue DapA protein. (A) Sequence of DapA with the N-terminal half denoted in green and the C-terminal half denoted in blue. Ligation sites are underlined and shown in red. (B) Successful synthetic strategy of DapA following the same coloring scheme as above. Image from (Weinstock 2014).

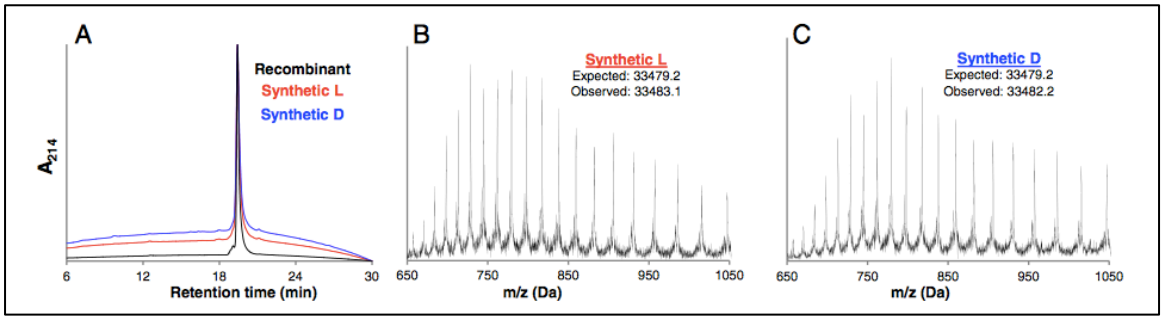


Figure 6: Analysis of DapA. (A) Analytical reverse-phase HPLC traces of recombinant, synthetic L-, and synthetic D-DapA. (B/C) Mass spec data obtained for synthetic L-DapA (B) and D-DapA (C). Image from (Weinstock 2014).

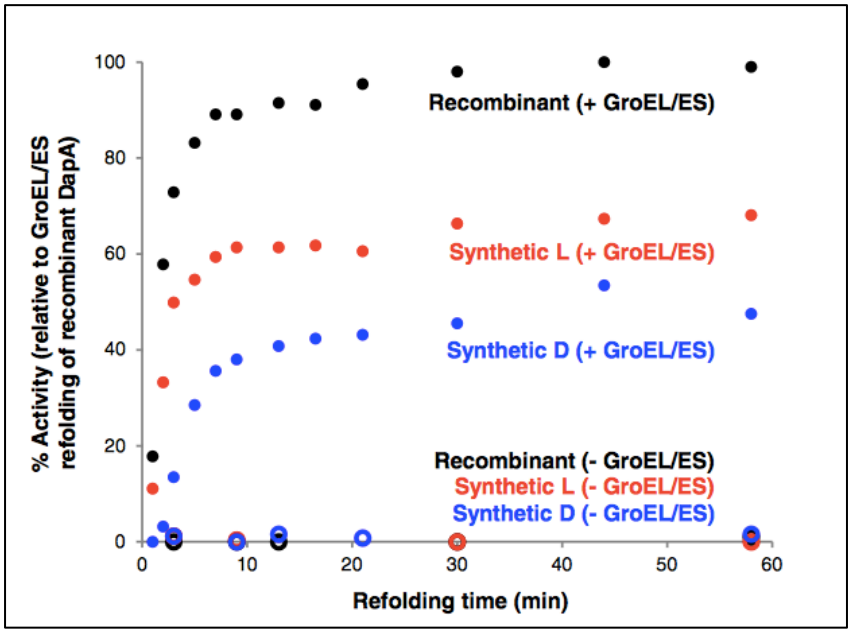


Figure 7: Analysis of GroEL/ES-mediated refolding of synthetic and recombinant DapA. Percent activity is normalized to recombinant (+ GroEL/ES). Synthetic L is shown in red and synthetic D is shown in blue. The open circles for all three samples observed at the bottom of the graph represent the refolding when GroEL/ES is absent. Image from (Weinstock 2014).

CHAPTER 2

DPO4 ACTIVITY ASSAYS AND SYNTHETIC PLAN

Dpo4 Background

Dpo4 (DNA polymerase IV) was discovered in thermophilic Archea in the genus Sulfolobus (Boudsocq 2001). It is a member of the Y-family polymerases known for being low-fidelity polymerases (Wu 2011). Y-family DNA polymerases are found throughout all kingdoms of life (Boudsocq 2004). The main function of these polymerases is to help cells deal with DNA damage, as they are capable of performing translesion synthesis (Boudsocq 2004). When DNA polymerases from other families encounter common types of DNA damage such as thymine-dimers or abasic sites, they stall and are unable to continue synthesizing DNA (McDonald 2006). Y-family polymerases, such as Dpo4, are able to rescue this process by synthesizing past these DNA lesions, allowing the stalled polymerase to complete its job. Because of this translesion synthesis capability, Y-family polymerases are of particular interest for use in attempting to replicate damaged ancient DNA (McDonald 2006).

There are multiple crystal structures of Dpo4, and through structure analysis it is apparent that there are four distinct domains (little finger, finger, palm, and thumb) as shown in Figure 8. In general, Y-family polymerases have been shown to have more flexible active sites that are more solvent accessible

than high-fidelity polymerases (Kirouac 2011). It is thought that because Dpo4 has a solvent-accessible active site, it is more permissive with dNTP incorporation. This property would explain the higher error rate and the ability to perform translesion synthesis (Kirouac 2011). Another interesting feature of Dpo4 is that it is highly selective for dNTPs over NTPs. However, with a single mutation (Y12A), Dpo4 gains the ability to incorporate NTPs, thus becoming a crude RNA polymerase (Kirouac 2011). The bulky tyrosine residue at position 12 is thought to clash with the 2'OH group of NTPs, preventing NTP incorporation through a mechanism dubbed the "steric gate" (Kirouac 2011).

We have chosen Dpo4 as a synthetic target for multiple reasons. First, its size of 352 residues is attractive to us as synthetic protein chemists because it is the smallest thermostable DNA polymerase on record, but would still be the largest synthetic protein on record. Most DNA polymerases are much larger and therefore outside of the range of current or near-future synthetic techniques. For example, the common bacterial DNA Polymerase II has 783 residues (small for a DNA polymerase). We also chose Dpo4 because of its ability to be both a DNA polymerase and a RNA polymerase with a single mutation at Y12A. Finally, we plan to synthesize Dpo4 in both L- and D- forms in order to enable mirror-image PCR, which could be useful in searching for D-life and also as a key intermediate towards generating a recombinant method for producing D-proteins via the *D. coli* project.

Synthesis Plan

In order to synthesize Dpo4, a protein containing 361 residues (including N-terminal His-tag), it was necessary to carefully analyze the sequence in order to formulate a synthesis plan. In the initial analysis, we identified potential ligation sites, Cys or Ala residues, and then selected the initial synthetic peptide segments. We also searched for any obvious mutations that would not harm the structure or function of Dpo4, but could greatly simplify the synthesis.

Our original synthesis plan divided Dpo4 into nine peptide segments (ranging from 26 to 53 residues). The number of starting peptide segments correlates to the number of handling steps (purifications, ligations, etc.) required to assemble the full-length protein. Our initial plan using nine peptide segments required at least 17 handling steps. Since each handling step results in a significant loss of material, we reduced the number of starting peptide segments from nine to six by synthesizing longer initial peptides. Figure 9 provides details on the six initial peptides, while the supplementary material presents data demonstrating the ability to synthesize and purify these longer segments.

Mutant Evaluation

After initial scouting syntheses, purifications, and ligations, we found three positions where mutations would greatly simplify our chemical synthesis efforts. In order to make these mutations, careful analysis is needed to minimize the risk that these changes will harm polymerase activity. When selecting candidate residues to mutate, we look at BLAST searches of homologous proteins to determine which residues are most common at the position of interest. We also

look at multiple crystal structures to see if there is anything obvious in the structure that would tolerate or even favor specific mutations.

After deciding on possible mutations, we expressed these mutated proteins. We then performed activity testing on these mutants and compared them to wild-type protein to determine if these mutations had a negative impact on the function of Dpo4. Our primary goal for the activity testing aspect of this project was to use a PCR reaction to validate the function of Dpo4. However, we initially struggled to get PCR to work with our recombinant enzyme. Knowing that PCR was our final goal, we designed a primer extension assay (see Methods) that allowed us to determine differences in thermostability between different mutants while we were troubleshooting the PCR assay. Using this primer extension assay we were able to measure the functionality and thermostability of different mutants compared to wild-type enzyme by heating the enzymes at different temperatures for 15 min followed by an extension assay (see Figures 10 and 11).

Cys 31

There is only one naturally occurring Cys residue in Dpo4 (at position 31). In order to complete the synthesis with this Cys intact, it would need to be protected with an ACM group during desulfurization of non-native Cys residues. Unfortunately, we have found ACM removal to be a highly challenging procedure in long peptide intermediates, commonly resulting in >50% loss of product. Further, ACM removal is not always 100% complete, requiring HPLC separation of our 361-residue product +/- ACM (likely very difficult). Two candidate residues

appealed to us for this mutation (Ser and Phe) based on BLAST searches and structural analysis. These mutations were generated and expressed recombinantly, and then tested for their impact on the thermostability of Dpo4.

Figure 10 directly compares wild-type to C31S and C31F using our primer extension assay. C31S is slightly more thermostable than C31F. C31S was fully active after 15 min heat treatment at 83.8° C, while C31F was not. C31S was not active after heat treatment at 85.4° C for 15 min, in contrast to wild-type, which survives heat treatment up to 86.3° C. Based on these data, we selected C31S to pursue further, but we still were concerned about the loss of thermostability in C31S compared to wild-type.

Ile 101

The ligation site between peptide segments 3 and 4 (see Figure 9) employs an Ile thioester based on the wild-type sequence. Unfortunately, isoleucine residues are very slow in NCL reactions (Figure 12) (Hackeng 1999) and after multiple attempts at 20-24 hour ligations with extremely poor yield, we decided to mutate this position to increase the reactivity of the thioester. Two potential changes that would provide the greatest benefit for the NCL reaction and are favorable via BLAST and structural analysis are Phe and Met. We made both of these mutants and performed activity testing to determine their effect on the functionality and thermostability of Dpo4 (Figure 11).

Figure 11 shows the primer extension results of our I101 mutants with I101M showing robust activity after heat treatment at up to 85.4° and reduced

activity after treatment at 86.3°. I101M was more thermostable than I101F, and I101M appears to be only slightly less thermostable than wild-type.

A155C

In our current synthesis strategy, the last step of the assembly requires desulfurization of Cys 155 to Ala. If this final desulfurization reaction is not 100% complete, then an HPLC purification would be necessary and require the separation of 361-residue Dpo4 with Cys from Dpo4 with Ala at position 155. Given the size of Dpo4 (~42 kD), separating these products with such a subtle difference (32 daltons with minor change in hydrophobicity) will likely be impossible. However, if we introduce the A155C mutation, we would shift the final step of the assembly to be a ligation reaction, which would result in a much easier final purification (resolving the complete protein from the N- and C-terminal halves). We are currently in the process of making the A155C mutation and will perform the activity testing in the near future.

PCR Activity Testing

As stated above, we initially tried to use our recombinant Dpo4 to perform PCR, but with little success. However, after we developed an in-depth purification and refolding protocol (see Methods), we observed small amounts of PCR product. With this new glimmer of hope, we spent a lot of time and effort optimizing this method and were eventually successful in creating a reproducible PCR assay (see Methods) that is more robust than our initial primer extension assay. During the optimization of the PCR assay, we struggled to obtain

reproducible data and determined that the heat sensor on our thermocycler had been malfunctioning. We tried many other PCR machines but were not able to set the lid temperature to less than 100° C, which proved to be a problem even though the heating block was set to the mid-80s for the denaturation step. Finally, we found a thermocycler in the Bass lab (Biorad C100 Touch Thermal Cycler) that had a fully customizable hot lid, and we were able to set the lid temperature to 70° C. This PCR method utilizes the gradient function on the thermocycler to assess the thermostability of proteins during the denaturation step of PCR. Figure 13 compares wild-type, C31S, and I101M at four different denaturation temperatures using a 15 or 5 s denaturation step at the listed temperatures.

These results confirmed the previous primer extension data with C31S unable to tolerate 15 s denaturation at any of the measured temperatures. However, we were able to rescue some C31S function using a 5 s denaturation protocol, and we observed good product yield after denaturation at 84.9° C, but not at 86° C. For reference, wild-type enzyme is fully active at 87.2° for both 5 and 15 s denaturations. I101M looks to be slightly less thermostable than wild-type, confirming the results seen in primer extension assays, but when the denaturation time was decreased to 5 s, it performs similarly to wild-type.

Knowing that we had found conditions where C31S and I101M were capable of performing PCR, we wanted to see if the double mutant (C31S & I101M) was also capable of performing PCR under these same conditions (Figure 14). Unfortunately the double mutant was unable to perform this PCR reaction. Based on our previous primer extension and PCR data, we knew that

C31S was much less thermostable than I101M. The inability of the double mutant to perform PCR is likely an additive effect of the two mutants, with C31S more of a contributing factor. This disappointing result forced us to reconsider the C31S mutation. Thus, we decided to generate a C31A construct, which was the next mutation on our list for C31.

Since we have such a narrow window for PCR with our recombinant enzymes, we developed alternate strategies to make the PCR reaction more robust. One of these ideas was to use a short linear template instead of a large plasmid template (7222 bp). We designed primers to amplify a ~370bp region of this larger plasmid with roughly 50 bp of excess DNA on either side of our PCR primers. Using this simplified template (sTemplate), we tested our wild-type recombinant enzyme against a commercial Dpo4 enzyme from New England Biolabs (NEB). Surprisingly, our wild-type enzyme was incapable of PCR while the commercial enzyme showed a very robust product (Figure 15). We hypothesized that the N-terminal His-tag on our construct may be interfering with our recombinant enzyme's performance and set out to design a construct with a TEV cleavage site for cleaving the His-tag after purification.

With two new constructs in hand (C31A and TEV-cleaved WT) we performed another round of activity testing. First, we wanted to see whether C31A outperformed C31S in terms of thermostability. Figure 16 confirms that C31A is the best option for mutation at this location as it performs the same as wild-type in this assay (5 s denaturation).

We also wanted to test whether TEV-cleaved wild-type was capable of performing similarly to the commercial NEB enzyme (i.e. at higher temperatures

and the sTemplate reaction). Figure 16 confirms that TEV-cleaved wild-type (denoted as NWT=naked wild-type) is a viable option moving forward, though this assay was not harsh enough to determine whether it performs similarly to the commercial enzyme.

We then decided to push the limits of our NWT enzyme by using higher temperatures for longer periods of time in the denaturation step and also to see whether NWT was capable of performing the sTemplate reaction. Figure 17 shows that NWT performs similarly to the commercial enzyme at higher temperatures for longer periods of time and that NWT is capable of performing the sTemplate PCR reaction. Based on this result, we conclude that the N-terminal His-tag hampers polymerase activity, thus we have found a solution to make our recombinant enzyme more robust.

Conclusions and Future Directions

In conclusion, we have identified three mutants that will help with the synthesis of Dpo4 and have validated C31A and I101M individually using a PCR assay. We also have discovered that the N-terminal His-tag that we have been using to purify our recombinant proteins and that also helps with the solubility of peptide segment 1-2 has a negative effect on the overall activity of the polymerase. These results have led to a preliminarily revised Dpo4 synthesis plan, but there are still recombinant activity assay experiments that need to be performed before we can finalize our synthesis plan.

First, we need to validate the double mutant C31A and I101M to confirm that these mutations do not impair the function of Dpo4 more than either of the

single mutants. Next, we need to confirm that the A155C mutation does not greatly affect the performance of the polymerase. We also need to validate that the proposed triple mutant (C31A, I101M, and A155C) is capable of performing PCR. Based on the results of these activity assays, we will decide the fate of the His-tag in synthetic Dpo4. If the double and triple mutants do not show robust activity with the His-tag and cleaving the tag increases their performance in PCR, then we will devise a synthesis plan that does not include the His-tag (either leave it out of the initial synthesis of peptide segment 1-2 or synthesize a TEV site that can be used to cleave the His-tag). After validating the multimutants and determining the best course of action regarding the His-tag, we can proceed with the chemical synthesis of Dpo4 in both L- and D- forms.

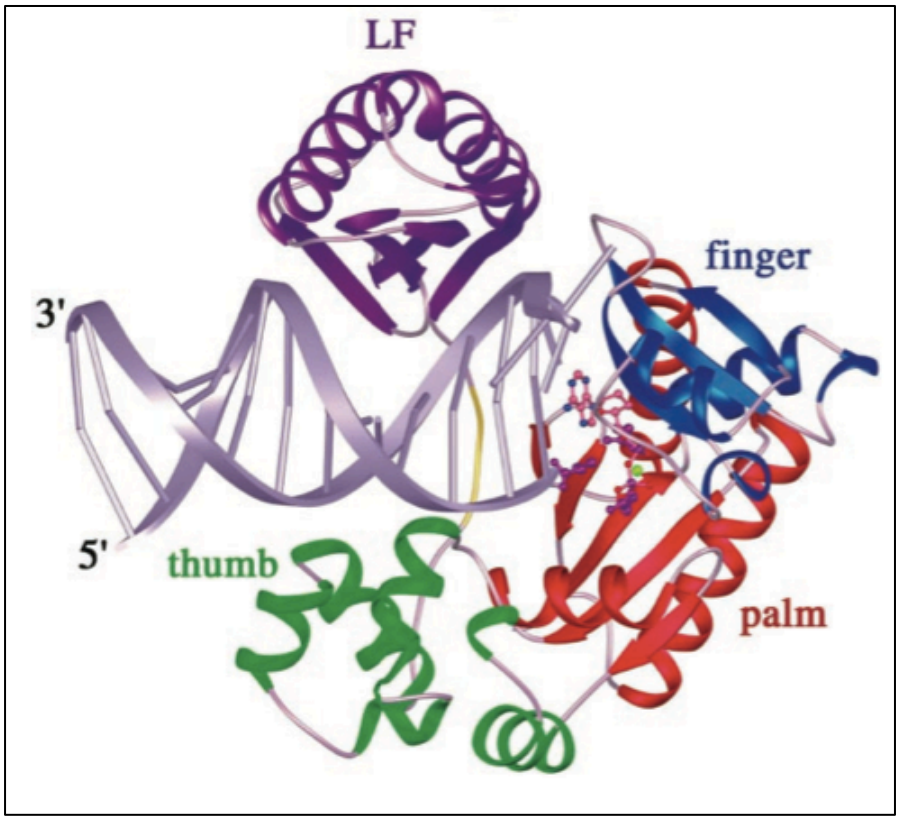


Figure 8: A ribbon diagram of Dpo4 in complex with DNA. This structure shows the different domains (little finger, finger, palm, and thumb) and how they interact with DNA. Image from (Boudsocq 2004).

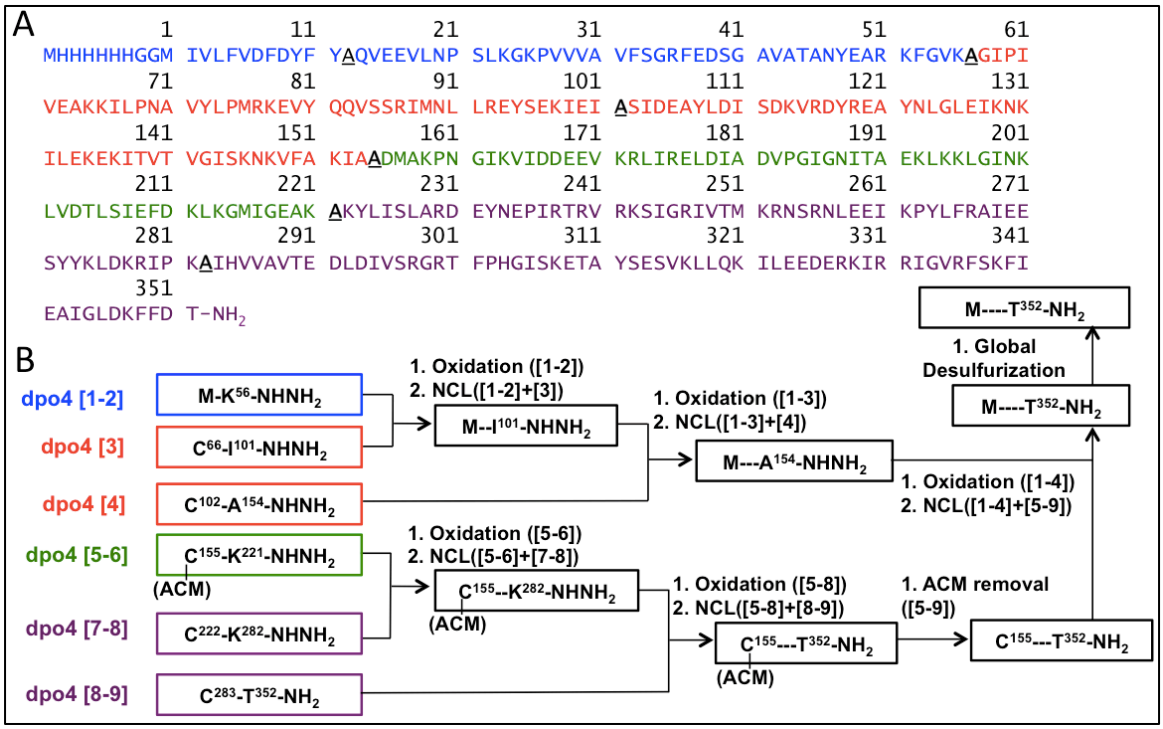


Figure 9: Chemical synthesis of Dpo4. (A) Target sequence for Dpo4. Finger domain (blue), palm domain (red), thumb domain (green), and the little finger domain (purple) are shown. Ligation sites are bold and underlined. (B) Final synthetic strategy. Peptide segments shown with the same color code as in A. Ligation thioesters and cysteine residues indicated on the initial peptides. Oxidation, native chemical ligation, ACM deprotection, and desulfurization steps are also shown.

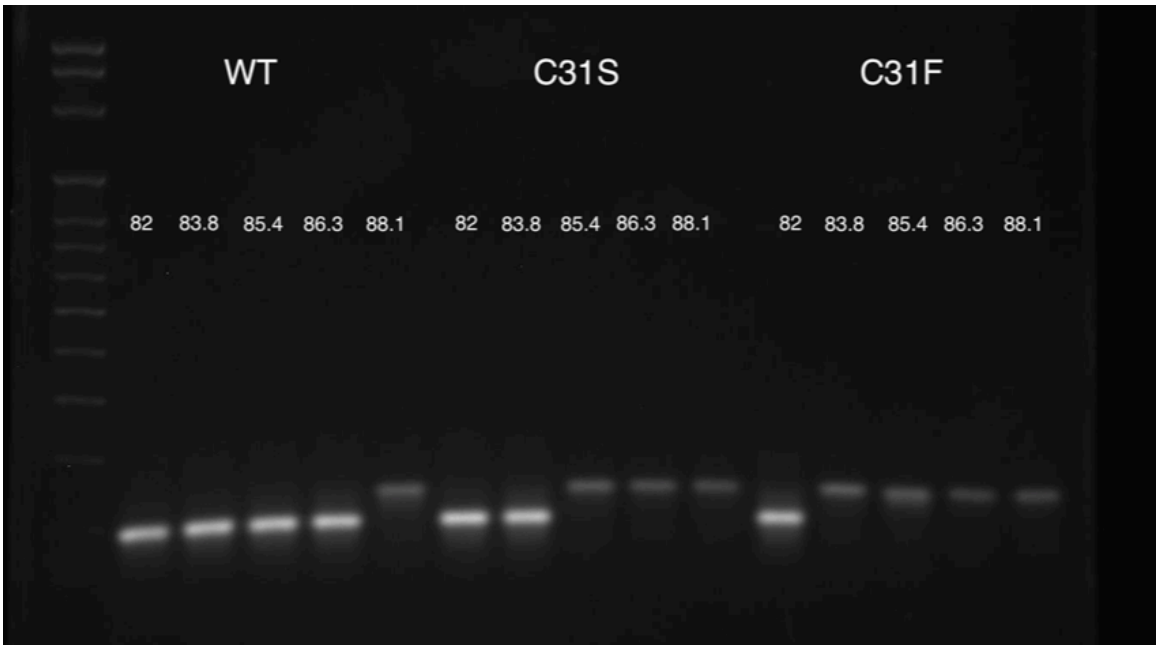


Figure 10: C31 mutants primer extension data. C31 mutants compared to wild-type. The figure is labeled by the mutant and also the 15-min incubation temperature in C before performing the primer extension assay. This assay allows us to see differences in thermostabilty compared to wild-type protein. The brighter bottom band is the fully extended 90-mer duplex while the upper band is the single stranded 90 bp oligo paired with a 20-mer primer (not extended). C31S is slightly more thermostable than C31F, but it is concerning that both of these mutants have lost substantial thermostability when compared to wild-type.

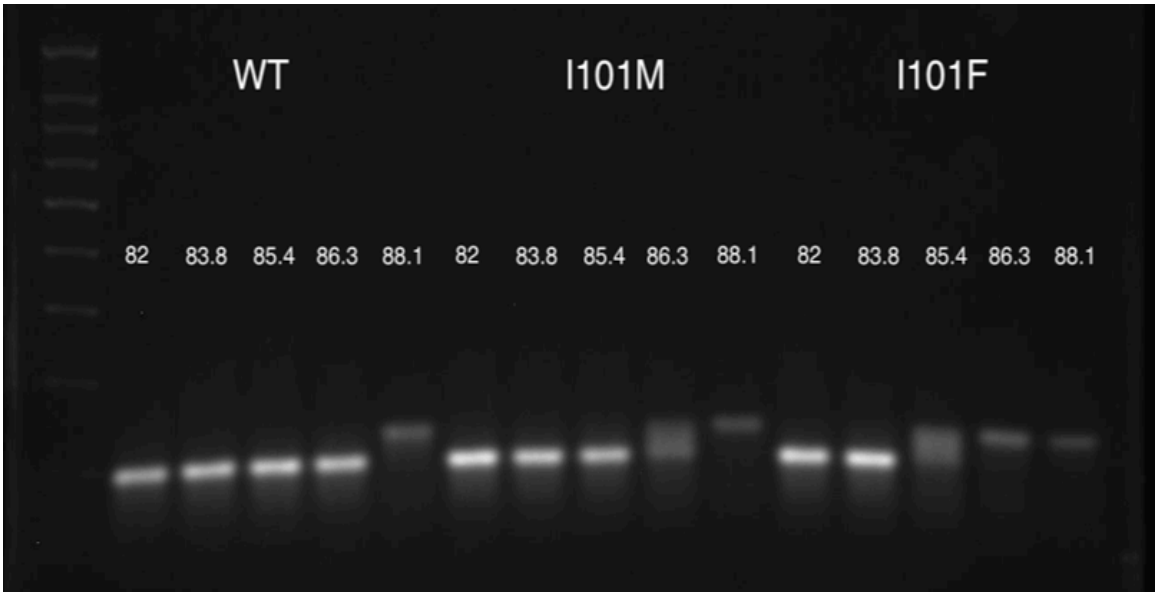


Figure 11: Primer extension data for I101 mutants. Enzyme samples were heated to the indicated temperatures in C before performing the primer extension assay. I101M is more thermostable than I101F. I101M is only slightly less thermostable than wild-type Dpo4.

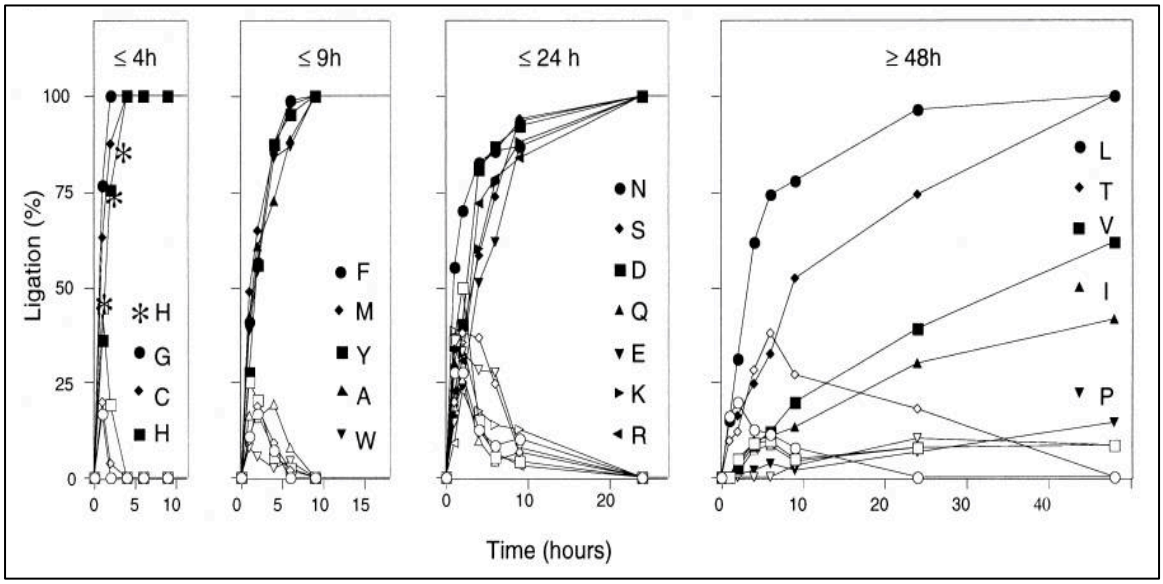


Figure 12: Kinetics of NCL reactions (different C-terminal thioesters). All ligations were done using the sequences LYRAX-to-CRANK with X representing all 20 amino acids shown in the graphs above. This figure clearly shows the importance of wisely choosing C-terminal thioesters in synthetic protein assembly projects. Image from (Hackeng 1999).

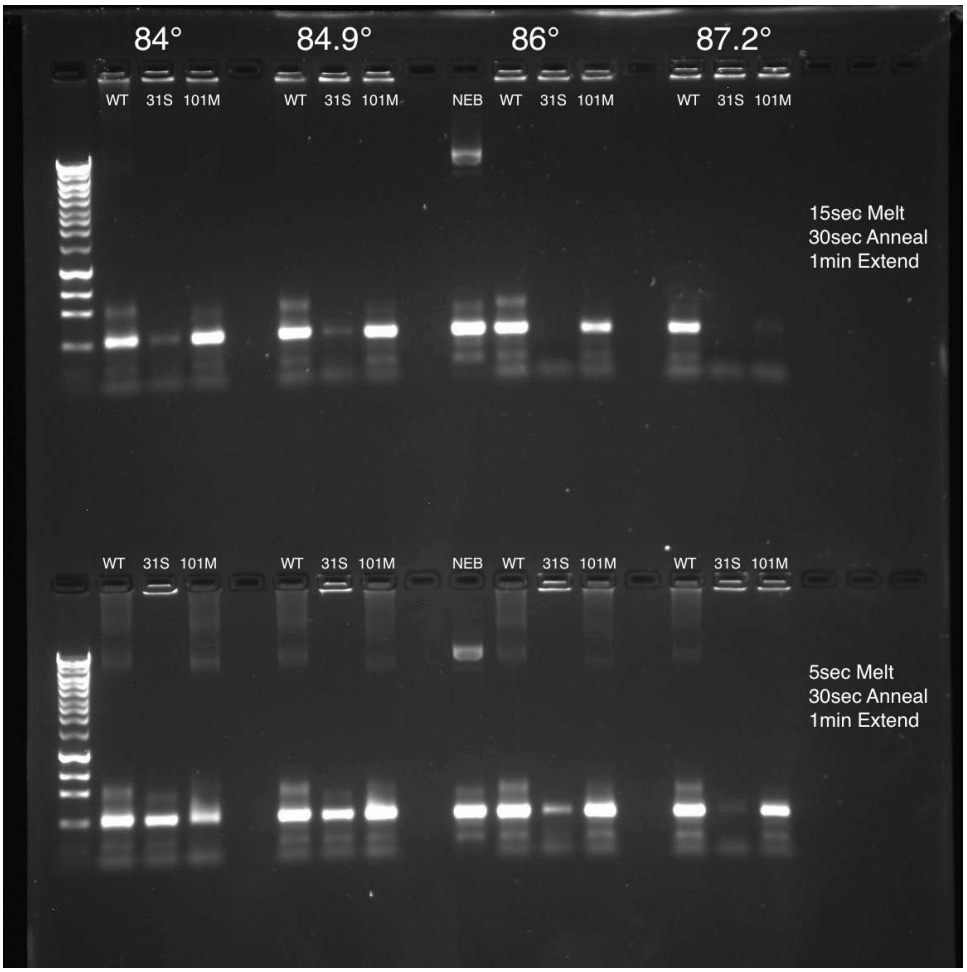


Figure 13: PCR assay data for wild-type, C31S, and I101M constructs. The top row underwent 15s denaturations at the listed temperatures while the bottom row underwent 5s denaturations at the same temperatures. (Top) Wild-type produces positive results at all shown temperatures. C31S produces very faint bands at the two lower temperatures with no activity at the higher two temperatures. I101M is active up until 87.2°. (Bottom) Wild-type and I101M perform similarly under the 5s denaturation conditions, and C31S yields robust results at the lower temperatures.

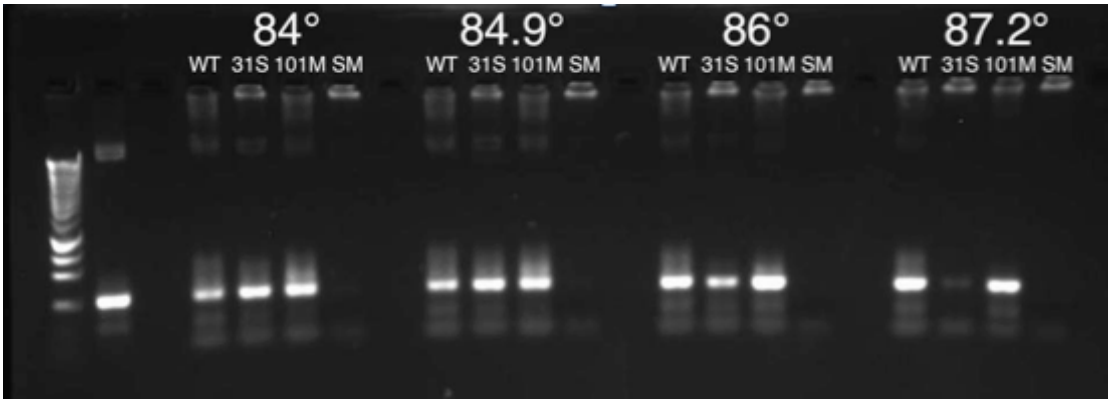


Figure 14: PCR assay data for wild-type, C31S, I101M, and the double mutant C31S/I101M. The three previously tested enzymes performed as expected, but the double mutant failed to produce positive results at any temperature.

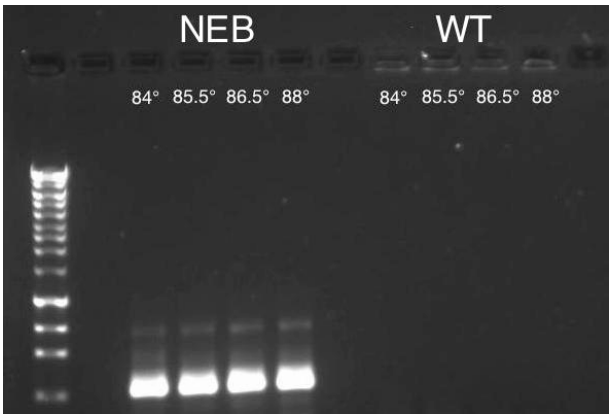


Figure 15: PCR assay data (linear simplified template (sTemp)). The commercial enzyme from New England Biolabs (NEB) produced very robust PCR product, while the His-tagged wild-type enzyme was inactive.

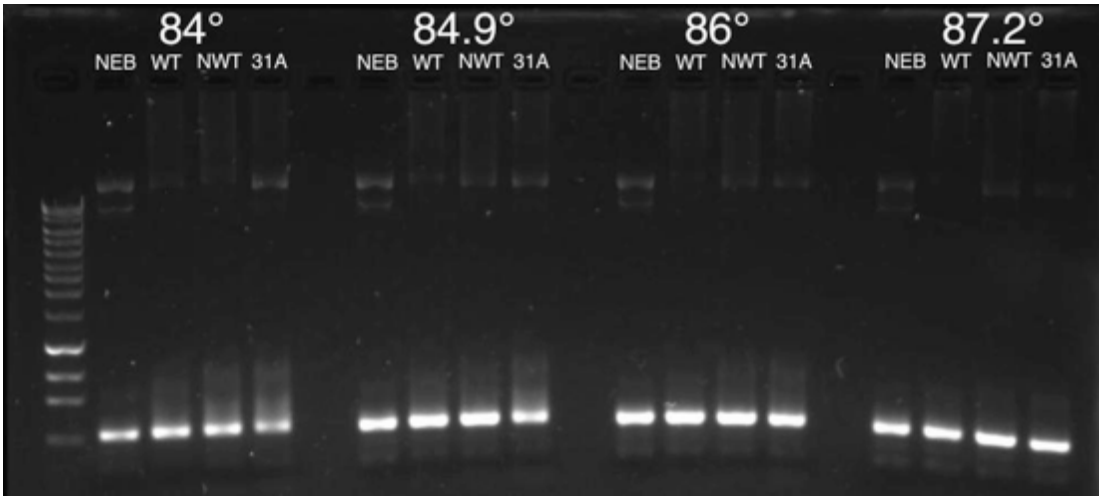


Figure 16: PCR assay data (C31A and TEV cleaved wild-type). TEV cleaved wild-type denoted as NWT (naked wild-type). C31A is capable of PCR at all of the listed temperatures and performs similarly to wild-type under the 5s denaturation protocol. NWT also performs similarly to wild-type under these conditions.

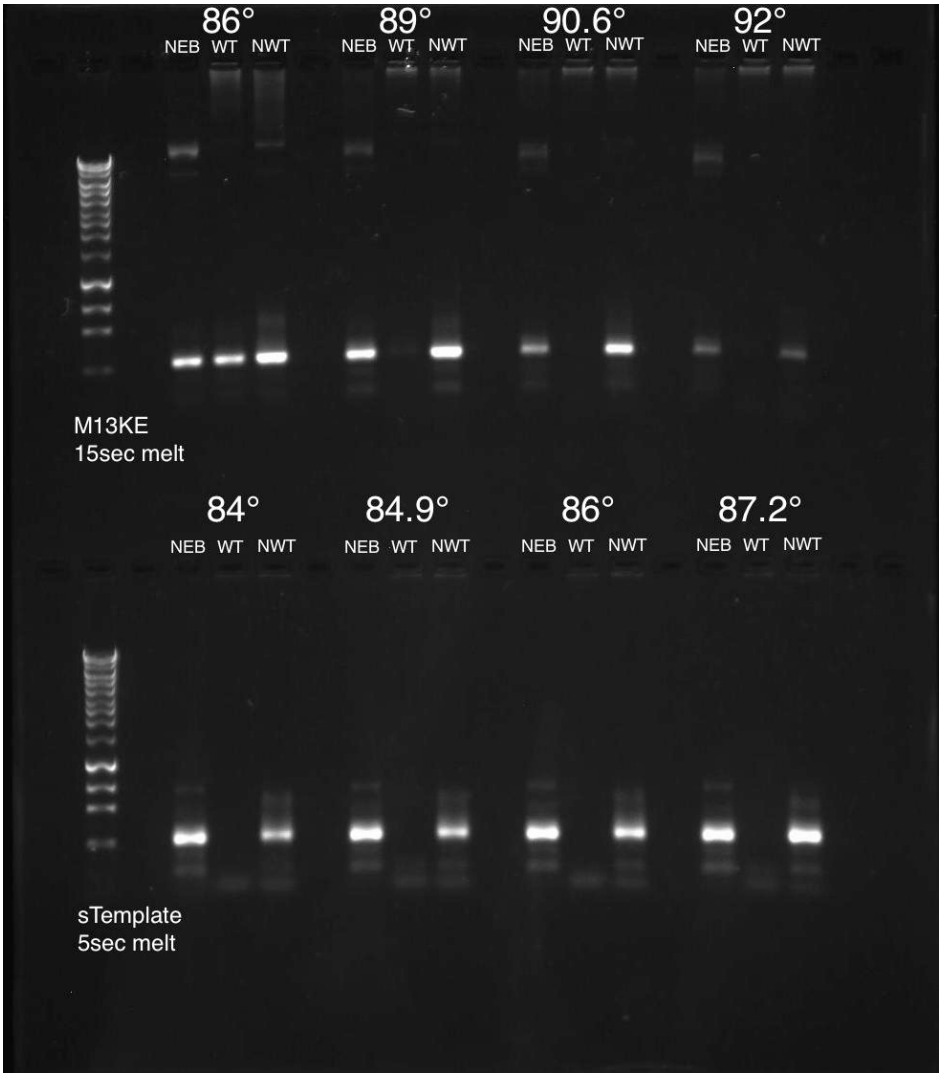


Figure 17: PCR assay data comparing NWT to commercial NEB enzyme. (Top) Higher temperatures for a longer period of time (15s). NWT performs similarly to NEB and both of these enzymes out-perform His-tagged wild-type. (Bottom) sTemplate reaction comparing His-tagged wild-type, NWT, and NEB. NEB and NWT enzymes are capable of performing the sTemplate reaction while His-tagged wild-type is not. The His-tag appears to negatively affect Dpo4 activity.

CHAPTER 3

METHODS

Cloning

Wild-type Dpo4 construct was cloned into pEXP5-CT vector used for T7-based, high-level expression of recombinant proteins by our collaborators in the George Church lab at Harvard. Mutants were made by quick-change or around-the-horn PCR methodology. Multimutant constructs were generated using restriction enzyme digests, gel purification, and ligation to yield double or triple mutants.

Expression and Purification

Dpo4 was expressed using BL21(DE3) pLysS competent cells in auto-induction (ZY) media. 2 mL of 1 M $MgSO_4$, 2 mL of 1000x metals mix, 2 mL of 1000x Ampicillin stock, 40 mL of 50x 5052, and 100 mL of 20x NPS were added to 1.856 L of ZY media. 500 mL of media was aliquoted into 2.8 L baffled bottom flasks and each culture was inoculated using a single colony. Cultures were grown at 37° C for ~16 h and harvested by centrifuge for 15 min at 5000 rpm (Beckman JLA-8.1000 rotor). Cell pellets were sonicated in Ni Binding Buffer (20 mM $NaPO_4$ pH 8, 300 mM NaCl, 10 mM imidazole) with the addition of lysozyme. The sonicated material was then spun at 15,000 rpm (Thermo Scientific FiberLite F18-12x50 rotor) for 30 min to separate the soluble protein fraction from the

insoluble fraction. The soluble fraction was adsorbed to Ni-NTA beads for 1 h and then eluted from the beads using elution buffer (20 mM NaPO4 pH 8, 300 mM NaCl, 250 mM imidazole). SDS-PAGE was used to confirm protein expression. The eluted proteins were then dialyzed overnight at 4° C into storage buffer (50 mM Tris pH 8, 50 mM NaCl, 1 mM DTT), acidified using 5% AcOH to pH 3, HPLC purified on a Phenomenex C4 column (10 µ, 300 Å, 10 x 250 mm), and lyophilized.

Refolding

Lyophilized Dpo4 powder was resuspended in denaturation buffer (6 M GuHCl, 50 mM Tris pH 8, 50 mM NaCl) to a concentration of ~25 µM in 1.5 mL final volume. The unfolded protein solution was then dialyzed back into storage buffer (50 mM Tris pH 8, 50 mM NaCl, 1 mM DTT) using Thermo Scientific Slide-a-Lyzer dialysis cassettes 3,500 MWCO overnight at room temperature in order to refold the protein. Following refolding, the sample was then spun on a table-top centrifuge at max speed for 10 min (to remove any particulates) and protein concentration was measured via A280 on a Biophotometer (extinction coefficient of 19,200 $cm^{-1} M^{-1}$ and light scattering correction of $A280_{corrected}$=A280-(1.7*A320)). This spun and refolded protein solution was used to perform activity assays described below.

Dpo4 Activity Assay-Primer Extension

Template/Primer Annealing

- Resuspend NEB-Lib-Extension and Ebo-fixed-SLLSA to a final concentration of 100 pmol/µL

Ebo-Fixed-SLLSA

AAAAAAAAAAAAAAAAGGTACCTTTCTATTCTCATTCTTCTCTTCTG
TGGCATTCTCTTTATTAAGTTTTCTCCGGTGGGGGTTCGGCCG

NEB-Lib-Extension

CTCTCTCGGCCGAACCCCCACC

Prepare Annealing Reaction

- Add 2 µL of Ebo-fixed-SLLSA to a PCR tube
- Add 4 µL of NEB-Lib-Extension to the PCR tube
- Add 44 µL 1X TE to a final volume of 50 µL
- Heat in thermocycler to 90 degrees for 5 min then turn machine off and let the annealing reactions cool slowly (~15min)

Gradient Enzyme Preheat

- Prepare 800 nM stocks of enzymes to be tested in NEB Thermopol buffer
- Run the gradient program with enzyme samples in the appropriate lanes (ususally between 79 and 87 degrees for 15 min)
- Cool the samples to room temperature on bench

Extension Reaction

- In 500uL eppendorf tubes, prepare the following for each reaction
    - 53 µL of ddH2O
    - 15 µL of 10x Thermopol buffer
    - 50 µL annealing rxn
    - 30 µL of 2.5mM dNTPs
    - 50 µL of gradient cooked 800nM enzyme
- Incubate in thermocycler @60 degrees for 10 min then put rxns on ice
- Negative control has no enzyme added and the excess volume made up with ddH2O
- Positive control is addition of WT enzyme that has not been preheated

Analysis

- Run samples on Agarose or PAGE gel

Dpo4 PCR Protocol

For each 25 µL reaction add:

- 2.5 µL Thermopol 10x buffer
- 1 µL Forward primer (10µM) M13 1470
- 1 µL Reverse primer (10µM) -96 gIII sequencing primer from NEB
- 1 µL dNTPs (5mM)
- 140 ng Template DNA M13KE plasmid
- 2 µL 12.5µM recombinant enzyme or 2µL NEB enzyme diluted 6X
- ddH2O (Fill to 25µL)

Primers

M13 1470: cgcaactatcggtatcaagc
-96 gIII: ccctcatagttagcgtaacg

Prepare Annealing Reaction

- Turn heat block on to 95°
- Combine everything listed above (except enzyme) into an eppendorf tube labeled Master Mix
    - Prepare excess Master Mix to account for pipetting inaccuracies (e.g. to test 3 mutants at 4 temperatures, prepare enough for 15 reactions)
- Heat at 95° for ~3 min
    - Annealing reaction will become milky white due to the Triton X-100 detergent present in the NEB Thermopol buffer
- Remove the metal block from the machine and let cool to room temperature (20-30 min.)
    - Solution should become clear again at lower temperatures

Enzyme Addition

- Split annealing reaction into the same number of small eppendorf tubes as mutants that will be tested
    - For example: to test 3 enzymes at 4 temperatures, add 23x5=115 µl into 3 different tubes
- Add enzyme to 1µM (2 µL of 12.5 µM enzyme per rxn)
    - Mix thouroughly
- Aliquot 25 µL into thin walled PCR tubes

Thermocycler Program (Gradient)

Biorad C1000 TouchThermal Cycler

- Set Gradient for melt step at 84°-92°
  - Standard gradient: H=84°, G=84.9°, F=86°, and E=87.2°
- Set lid to 70°

Melt:	T=gradient	5 sec
Anneal:	T=56°	30 sec
Extend:	T=60°	1 min
25 Cycles		
Polish:	T=60°	5 min
Store:	T=4°	∞

Agarose Gel Electrophoresis

- Prepare a 1% agarose gel
- Load 10 µL sample using Affymetrix 5X RapidRun Loading Dye
- Run at 160 V for ~1hr

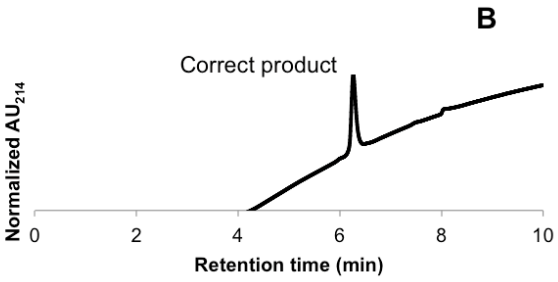
SUPPLEMENTARY MATERIAL

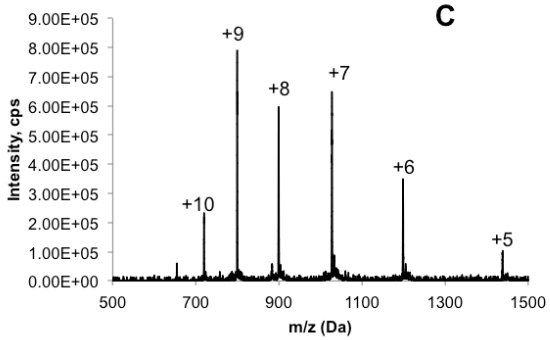
The following figures contain RP-HPLC and mass spec data for all purified starting peptides. (A) table showing calculated and observed mass for each peptide as well as the sequence and ions observed. (B) RP-HPLC trace on C4-Aeris column. (C) mass spec m/z readout. (D) deconvoluted total mass.

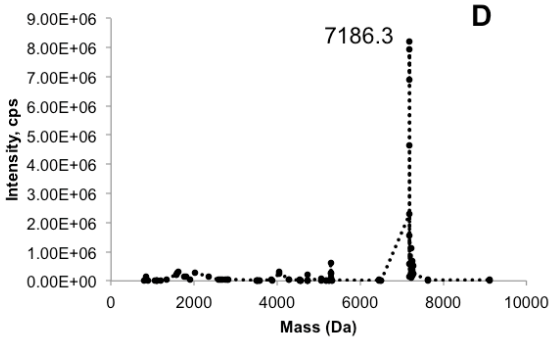
Dpo4 [1-2]

A

Peptide	Dpo4 [1-2]
Calculated Mass (Da)	7187.1
Observed Mass Avg (Da)	7186.3
Ions	599.8 (+12), 654.3 (+11), 719.6 (+10), 799.5 (+9), 899.3 (+8), 1027.6 (+7), 1198.7 (+6), 1437.8 (+5)
Sequence	MHHHHHHGGIVLFVDFDYFYAQVEEV LNPSLKGKPVVVSVFSGRFEDSGAV ATANYEARKFGVK-NHNH2



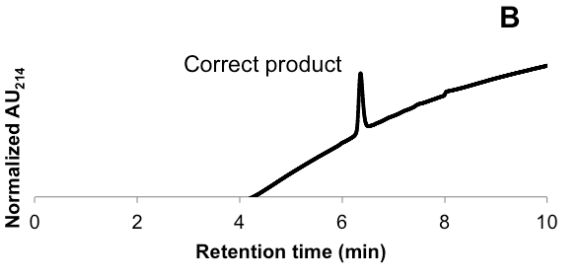


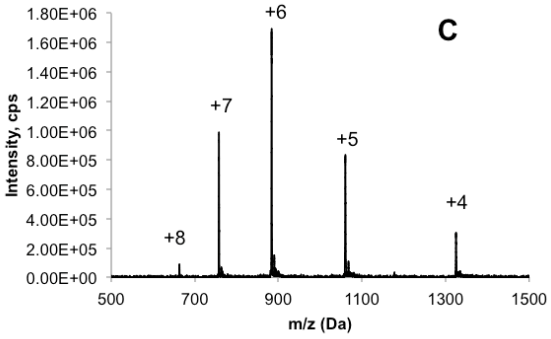


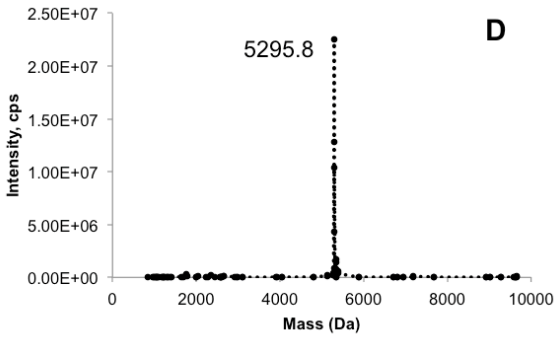
Dpo4 [3]

A

Peptide	Dpo4 [3]
Calculated Mass (Da)	5296.3
Observed Mass Avg (Da)	5295.8
Ions	662.9 (+8), 757.6 (+7), 883.6 (+6), 1060.1 (+5), 1324.9 (+4)
Sequence	CGIPIVEAKKILPNAVYLPMRKEVYQQ VSSRIMNLLREYSEKIEM-NHNH2



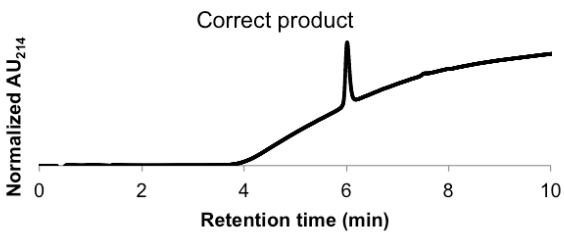


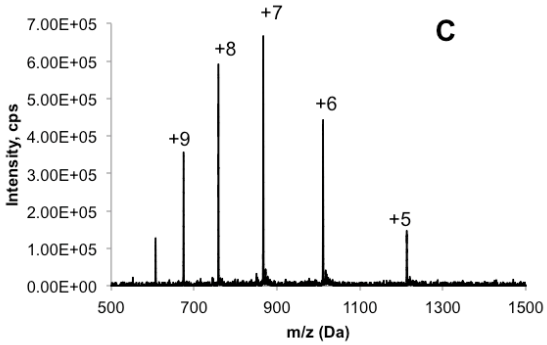


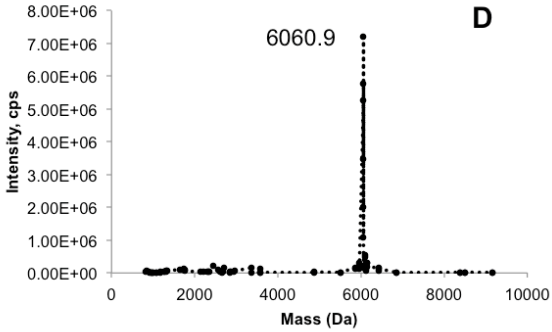
Dpo4 [4]

A

Peptide	Dpo4 [4]
Calculated Mass (Da)	6060.9
Observed Mass Avg (Da)	6060.9
Ions	552.1 (+11), 607.1 (+10), 674.5 (+9), 758.7 (+8), 866.9 (+7), 1011.1 (+6), 1213.1 (+5)
Sequence	CSIDEAYLDISDKVRDYREAYNLGLEIK NKILEKEKITVTVGISKNKVFAKIA-NHNH2



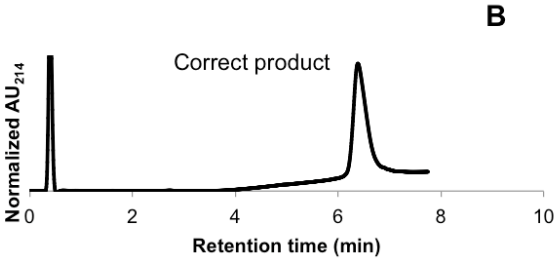


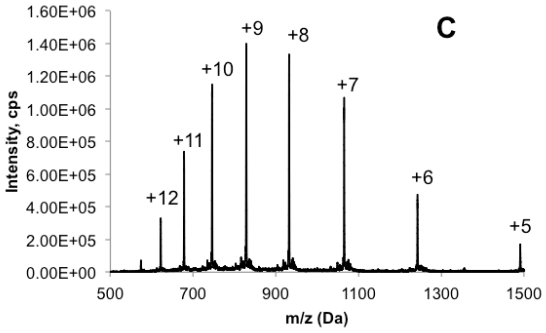


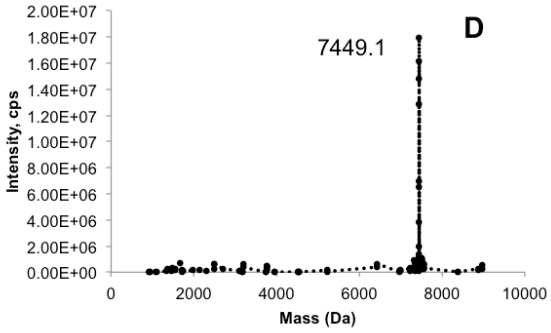
Dpo4 [5-6]

A

Peptide	Dpo4 [5-6]
Calculated Mass (Da)	7449.5
Observed Mass Avg (Da)	7449.1
Ions	574.0 (+13), 621.8 (+12), 678.3 (+11), 745.9 (+10), 828.6 (+9), 932.1 (+8), 1065.4 (+7), 1242.6 (+6), 1491.3 (+5)
Sequence	C(Acm)DMAKPNGIKVIDDEEVKRLIRELDIADVPGIGNITAEKLKKLGINKLVDTLSIEFDKLKGMIGEAK-NHNH2



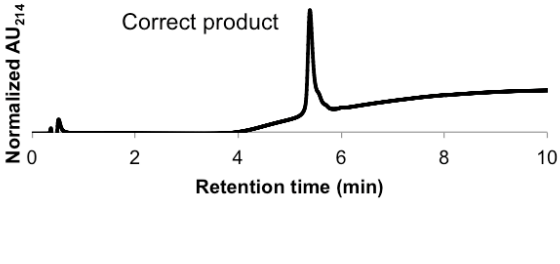


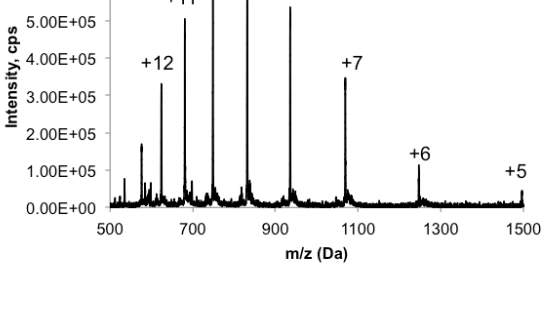


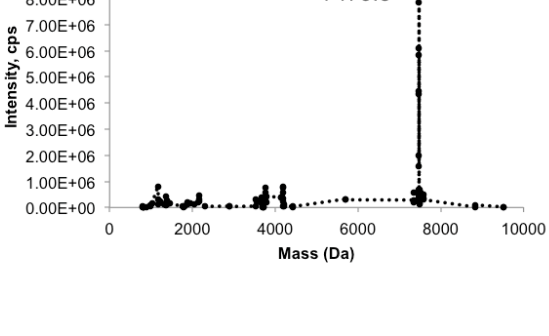
Dpo4 [7-8]

A

Peptide	Dpo4 [7-8]
Calculated Mass (Da)	7476.7
Observed Mass Avg (Da)	7475.8
Ions	534.9 (+14), 576.1 (+13), 623.9 (+12), 680.6 (+11), 748.6 (+10), 831.6 (+9), 935.6 (+8), 1069.2 (+7), 1247.0 (+6)
Sequence	CKYLISLARDEYNEPIRTRVRKSIGRIVTMKRNSRNLEEIKPYLFRAIEESYYKLDKRIPK-NHNH2



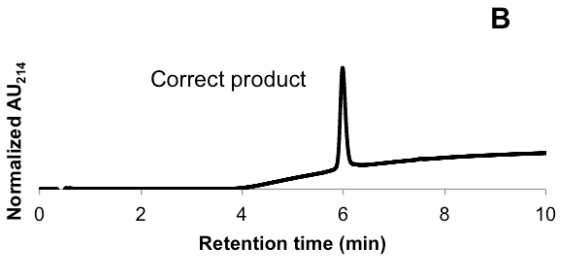


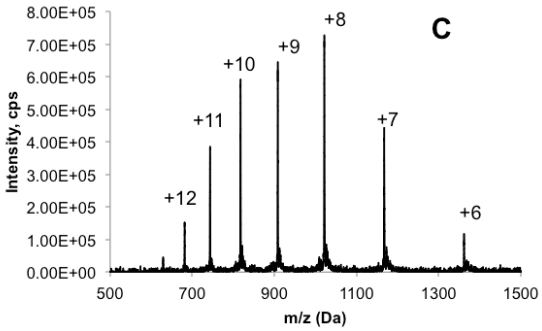


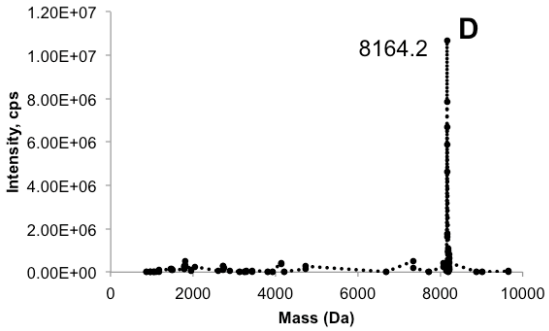
Dpo4 [8-9]

A

Peptide	Dpo4 [8-9]
Calculated Mass (Da)	8165.4
Observed Mass Avg (Da)	8164.2
Ions	681.3 (+12), 743.2 (+11), 817.5 (+10), 908.1 (+9), 1021.5 (+8), 1167.4 (+7), 1361.9 (+6)
	CIHVVAVTEDLDIVSRGRTFPHGISKE TAYSESVKLLQKILEEDERKIRRIGVRF SKFIEAIGLDKFFDTGG-NH2







REFERENCES

Boudsocq F, Kokoska RJ, Plosky BS, Vaisman A, Ling H, et al. Investigating the role of the little finger domain of Y-family DNA polymerases in low fidelity synthesis and translesion replication. J Biol Chem. 2004 Jul 30;279(31):32932-40.

Boudsocq F, Iwai S, Hanaoka F, Woodgate R. Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic poleta. Nucleic Acids Res. 2001 Nov 15;29(22):4607-16.

Carpino LA. 9-Fluorenylmethoxycarbonyl amino-protecting group. The Journal of Organic Chemistry. 1972; 37(22):3404-3409.

Clinton TR, Weinstock MT, Jacobsen MT, Szabo-Fresnais N, Pandya MJ, et al. Design and characterization of ebolavirus GP prehairpin intermediate mimics as drug targets. Protein Sci. 2014 Oct 6.

Dawson PE, Muir TW, Clark-Lewis I, Kent SB. Synthesis of proteins by native chemical ligation. Science. 1994 Nov 4;266(5186):776-9.

Dawson PE, Kent SB. Synthesis of native proteins by chemical ligation. Annu Rev Biochem. 2000;69:923-60.

Francis JN, Redman JS, Eckert DM, Kay MS. Design of a modular tetrameric scaffold for the synthesis of membrane-localized D-peptide inhibitors of HIV-1 entry. Bioconjug Chem. 2012 Jun 20;23(6):1252-8.

Frydman J. Folding of newly translated proteins in vivo: the role of molecular chaperones. Annu Rev Biochem. 2001;70:603-47.

Harris KM, Flemer S Jr, Hondal RJ. Studies on deprotection of cysteine and selenocysteine side-chain protecting groups. J Pept Sci. 2007 Feb;13(2):81-93.

Hackeng TM, Griffin JH, Dawson PE. Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. Proc Natl Acad Sci U S A. 1999 Aug 31;96(18):10068-73.

Isidro-Llobet A, Alvarez M, Albericio F. Amino acid-protecting groups. Chem Rev. 2009 Jun;109(6):2455-504.

Johnson EC, Kent SB. Insights into the mechanism and catalysis of the native chemical ligation reaction. J Am Chem Soc. 2006 May 24;128(20):6640-6.

Kent SB. Chemical synthesis of peptides and proteins. Annu Rev Biochem. 1988;57:957-89.

Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, et al. Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. Cell. 2005 Jul 29;122(2):209-20.

Kimmerlin T, Seebach D. '100 years of peptide synthesis': ligation methods for peptide and protein synthesis with applications to beta-peptide assemblies. J Pept Res. 2005 Feb;65(2):229-60.

Kirouac KN, Suo Z, Ling H. Structural mechanism of ribonucleotide discrimination by a Y-family DNA polymerase. J Mol Biol. 2011 Apr 1;407(3):382-90.

Malins LR, Mitchell NJ, Payne RJ. Peptide ligation chemistry at selenol amino acids. J Pept Sci. 2013 Nov 28.

Merrifield RB. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. Journal of the American Chemical Society. 1963; 85(14):2149-2154.

McDonald JP, Hall A, Gasparutto D, Cadet J, Ballantyne J, et al. Novel thermostable Y-family polymerases: applications for the PCR amplification of damaged or ancient DNAs. Nucleic Acids Res. 2006;34(4):1102-11.

Saïda F. Overview on the expression of toxic gene products in Escherichia coli. Curr Protoc Protein Sci. 2007 Nov;Chapter 5:Unit 5.19.

Weinstock MT, Francis JN, Redman JS, Kay MS. Protease-resistant peptide design-empowering nature's fragile warriors against HIV. Biopolymers. 2012;98(5):431-42.

Weinstock MT, Jacobsen MT, Kay MS. Synthesis and folding of a mirror-image enzyme reveals ambidextrous chaperone activity. Proc Natl Acad Sci U S A. 2014 Aug 12;111(32):11679-84.

Welch BD, Francis JN, Redman JS, Paul S, Weinstock MT, et al. Design of a potent D-peptide HIV-1 entry inhibitor with a strong barrier to resistance. J Virol. 2010 Nov;84(21):11235-44.

Wu Y, Wilson RC, Pata JD. The Y-family DNA polymerase Dpo4 uses a template slippage mechanism to create single-base deletions. J Bacteriol. 2011 May;193(10):2630-6.

Xie J, Schultz PG. Adding amino acids to the genetic repertoire. Curr Opin Chem Biol. 2005 Dec;9(6):548-54.

Yan LZ, Dawson PE. Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. J Am Chem Soc. 2001 Jan 31;123(4):526-33.